(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
31 December 2003 (31.12.2003)

PCT

(10) International Publication Number
WO 2004/001044 A1

(51) International Patent Classification[7]: C12N 15/10, 15/09, 15/11, 15/63

(21) International Application Number: PCT/SE2003/001077

(22) International Filing Date: 23 June 2003 (23.06.2003)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
60/390,108 21 June 2002 (21.06.2002) US

(71) Applicant *(for all designated States except US)*: **SINOGENOMAX COMPANY LTD.** [CN/CN]; #3-707 North Yongchang Road, BDA, Beijing 100176 (CN).

(72) Inventors; and

(75) Inventors/Applicants *(for US only)*: **LIANG, Zicai** [CN/SE]; Sevensvägen 4, S-174 46 Sundbyberg (SE). **ZHANG, Hong-Yan** [CN/SE]; Sevensvägen 4, S-174 46 Sundbyberg (SE). **CHEN, Meihong** [CN/CN]; North Yongchang Road, BDA, Beijing 100176 (CN). **SHEN, Yan** [CN/CN]; North Yongchang Road, BDA, Beijing 100176 (CN).

(74) Agent: **DR. LUDWIG BRANN PATENTBYRÅ AB**; P.O. Box 171 92, S-104 62 Stockholm (SE).

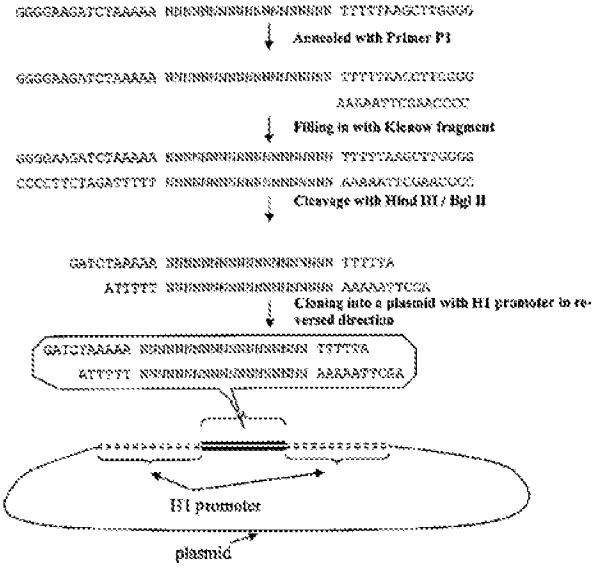
(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NI, NO, NZ, OM, PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, HU, IE, IT, LU, MC, NL, PT, RO, SE, SI, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG).

Declaration under Rule 4.17:
— *of inventorship (Rule 4.17(iv)) for US only*

[Continued on next page]

(54) Title: RANDOMISED DNA LIBRARIES AND DOUBLE-STRANDED RNA LIBRARIES, USE AND METHOD OF PRODUCTION THEREOF



(57) Abstract: This invention relates to DNA libraries based on plasmid or viral vectors that can express double-stranded RNA of 10-30 base pairs in length with all possible sequences, where each of the double stranded RNA is formed by a single RNA molecule in the form of hairpin, or formed by two separate RNA molecules with different 3'-overhangs. Each single member in such a DNA library encodes all components of a double stranded RNA as specified above. Such a library can be used in screening for double stranded RNA species that can induce a given phenotype without prior knowledge of their target genes. This invention further relates to a method to generate such a DNA library.

WO 2004/001044 A1

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

RANDOMISED DNA LIBRARIES AND DOUBLE-STRANDED RNA LIBRARIES, USE AND METHOD OF PRODUCTION THEREOF

TECHNICAL FIELD

This invention relates to DNA libraries based on plasmid or viral vectors that can express double-stranded RNA of 10-30 base pairs in length with all possible sequences, where each of the double stranded RNA is formed by a single RNA molecule in the form of hairpin, or formed by two separate RNA molecules with different 3'-overhangs. Each single member in such a DNA library encodes all components of a double stranded RNA as specified above. Such a library can be used in screening for double stranded RNA species that can induce a given phenotype without prior knowledge of their target genes. This invention further relates to a method to generate such a DNA library.

BACKGROUND ART

Messenger RNA (mRNA) is normally perceived as the information-carrying intermediate in protein synthesis that is transcribed by RNA polymerase from a DNA template and subsequently translated by ribosome to generate protein molecules. Recently more data have demonstrated that many genes are transcribed into RNA molecules that are not translated into proteins at all (Okazaki Y et. Al., *Nature*; 420(6915):563-573 (**2002**)). Some of the untranslated RNA were found to carry out functions in the regulation of the other mRNA by inducing the degradation of the mRNA in a sequence specific manner (Ambros V., *Cell*; 113(6):673-676 (**2003**)). This is in good agreement with the recent finding that double stranded RNA and synthetic siRNA can also induce cognate mRNA degradation in a wide range of organisms (McManus MT, Sharp PA., *Nature Rev Genet.*; 3(10):737-747 (**2002**)). Long double stranded RNA was found to induce intensive non-specific inhibition of RNA synthesis in mammalian cells, but siRNA can bypass this obstacle and still maintain the strong inhibitory effect on target gene which shares sequence identity with the siRNA (Elbashir SM et al., *Nature*; 411(6836):494-498 (**2001**)). This has made siRNA a primary tool for gene knockdown in functional genomics. SiRNA also has

the potential to become drugs that can be used to cure a disease by reducing the activity of disease related gene.

SiRNA are generally double stranded RNA of 19-25 base pairs that are either formed by a single RNA molecule in the form of hairpin or formed by two separate RNA molecules, with different 3'-overhangs. SiRNA can be produced in three ways: chemical synthesis; expression from DNA vectors under the drive of a promoter; and RNase III (Dicer) cleavage of long double stranded RNA. All siRNA that have been used so far are designed to target a segment of a predefined gene.

SUMMARY OF INVENTION

The present invention relates to DNA libraries, each of which contains all possible permutations (permutation refers to different sequences) of double-stranded RNA of certain length. Such DNA libraries can be easily configured to produce all permutations of siRNA. It provides a high throughput screening method for double stranded RNA (as well as siRNA) in a target-independent manner for indications related to any given phenotype. More specifically, the siRNA encoded by such libraries can be used in such screening either individually, or as a mixture of any complexity, without the burden of knowing its sequence or its target gene. This method can overcome two major obstacles in siRNA application: 1) the incomplete knowledge about the transcriptome of each organism. According to the recent data from mouse transcriptome analysis, our knowledge about the transcriptome of this best understood model animal is still far from complete. Much less is known about the transcriptome of human and other animals. Since the application of our library does not require any prior information about the target sequence, it will allow immediate implementation of genome-wide siRNA screening in any orgainisms. 2) the extraordinarily high cost of siRNA. No matter how the siRNA is prepared, the cost of making siRNA targeting all known mRNA of an organism is extremely high. A single regenerate-able DNA library that contains all permutation of siRNA that can be applied in any organisms virtually reduces the cost of siRNA production to a minimum level.

Accordingly, in one aspect the present invention relates to a DNA library for the production of a library of double stranded RNA molecules of a predefined length in the range of 10-30 base pairs in living cells, wherein the sequence(s) of the DNA region (or regions) encoding the double stranded part of double stranded RNA mole-

cule(s) is randomized in a number selected from 4 to all nucleotide positions, and wherein both strands of said double stranded RNA molecule is produced from a single member of the DNA library. The invention also provides a kit containing the DNA library.

In another aspect the present invention provides a method of preparing the DNA library.

In yet another aspect the invention relates to an RNA library obtained from the DNA library.

Further aspects and advantages of the invention will become evident hereinafter from the following detailed description and attached claims.

DESCRIPTION OF THE DRAWINGS

Figure 1 shows an example of construction of DNA library that can encode all permutations of double stranded RNA of a certain length. Example 1, a DNA library that can encode all double stranded RNA with 19 base pair duplex region and 3´ poly U over hangs. In Figure 1A, the cloning strategy is shown. In Figure 1B, experimental verification of the quality of the library is demonstrated. As shown in the agarose gel, single clone (1x), and pools of 10 clones (10x), and pools of 30 clones give rise to the a single expected band after enzyme cleavage, suggesting that most clones in the library contain the expected insert. The same procedure can be used to produce such DNA libraries encoding different length (10-30 base pair) of double stranded RNA, as well as such DNA libraries with only part of the DNA sequence (4-30 nt) randomized.

Figure 2 shows the construction of a plasmid to verify that the presence of two promoters and two terminators in opposite sides of the RNA coding region can afford efficient down-regulation of the expression of the target gene. With all scientific knowledge available as of today, such an efficient down regulation can only be achieved by the efficient production of double stranded RNA from the plasmid. Thus it is concluded that this plasmid can efficiently produced double stranded RNA in living cells. A shows the cloning strategy. B shows the gel analysis verified that the

designed fragment is inserted into the plasmid. C illustrates cell assay verified that the resulting plasmid induces efficient inhibition of target gene Renilla luciferase.

Figure 3 shows an example of an alternative method of generating DNA libraries that encode all permutations of double stranded RNA of a given length. In Figure A, the cloning strategy is shown. In Figure B, sequences of the different segments in A with key restriction sites underlined are shown. The same procedure can be used to produce such DNA libraries encoding different length (10-30 base pair) of double stranded RNA, as well as such DNA libraries with only part of the DNA sequence (4-30 nt) randomized.

Figure 4 shows another alternative method of generating DNA libraries that encode all permutations of double stranded RNA of a given length. A illustrates the cloning strategy. B illustrates sequences of the different segments in A with key restriction sites underlined. The same procedure can be used to produce such DNA libraries encoding different length (10-30 base pair) of double stranded RNA, as well as such DNA libraries with only part of the DNA sequence (4-30 nt) randomized.

DETAILED DESCRIPTION OF THE INVENTION

Small interference RNA (siRNA) is a term initially used to define short double stranded RNA that have a 19-21 nt double-stranded region nested between 3'-UU or TT or other single stranded overhangs. A number of variations of this original form of siRNA (such as hairpin-type) have been introduced lately. Such siRNA can be used to reduce the expression of genes having identical sequence to the siRNA double stranded region in cells from a variety of different organisms. While longer double stranded DNA and RNA also could be produced by means of the methods of the invention, the libraries of the invention have been restricted to double stranded DNA and RNA of a length of 10-30 base pairs, since above the length of 30 base pairs, the nucleotides will be more likely to produce an immunoresponse, and other disturbing side-effects when transfected into living cells.

SiRNA are initially chemically synthesized, but several methods have been introduced to generate siRNA enzymatically, using viral promoters such as T7 promoter, or microRNA promoter such as H1 or U6, in free form or in plasmid or viral vectors.

The current invention provides a method to construct DNA libraries encoding random siRNA libraries. Such a library differs from the prior art in that in the prior art, one would have to design the siRNA according to a known sequence of the gene, whereas from the present library one can screen through a fully random panel of different siRNA (without the need of prior knowledge of their sequences or their target sequences) to look for phenotypes associated with each siRNA, and then identify the genes related to each siRNA de novo.

Construction of DNA libraries containing a single randomized region

The challenge of making a fully randomized DNA library based on plasmids or viral vectors encoding all permutations of siRNA is to make sure that each member of the DNA library expresses a distinct and complete double stranded RNA. None of the existing methods of making vector-based siRNA (short double stranded RNA) can meet this challenge.

The current invention describes the construction of a random DNA library with only one randomized region. Then for each plasmid, two promoters will drive the transcription of this region from the opposite direction to produce the two complementary RNA strands separately. Two transcription terminators were placed at each end of the randomized region to make sure that RNA of a defined length can be produced from each direction. The advantage of this approach is to avoid the troublesome cloning procedure in the dual-region system as will be described beneath for creating two reverse complementing regions in each individual plasmid. One example of the promoters that can be used in such a system is the RNA polymerase III promoters H1 or U6. For RNA polymerase III, a stretch of TTTTT is needed for the proper termination of the transcription. In order to use this RNA polymerase to drive expression of the same region from both directions, the TTTTT stretch has to be inserted on the both ends of the randomized region. There is one problem though: the RNA polymerase III promoters has to be placed immediately next to the randomized region to ensure proper transcription start from the precise location of the beginning of the randomized region, but those promoters does not contain a AAAAA stretch that would allow the TTTTT terminator to appear on the opposite direction. The only way this can be done is to mutate the RNA polymerase III promoters to insert such a AAAAA stretch, and nobody knows how the insertion of the AAAAA stretch will affect the transcription starting, and the rate of transcription. As will be shown below, we mutated the H1 RNA polymerase III promoter and inserted an

AAAAA stretch at the end of the promoter and verified that the mutated promoter support proper transcription start and product of effective siRNA. Thus, we first started to construct a plasmid library with the termination signal placed on both sides of the randomized region (Figure 1).

Construction of the vector with dual-H1 promoter against renilla luciferase

A plasmid with two mutated RNA polymerase III promoters, each embedding one transcription terminator sequence for the other promoter, was constructed with the siRNA region designed to target a model molecule Renilla luciferase (Figure 2). The key finding that such a plasmid can support the successful production of effective siRNA duplex from a single target sequence of 19 bp forms the basis of constructing fully randomized siRNA library that have only one randomized region (Figure 2).

Mutation of the H 1 RNA polymerase III promoters and construction of the example plasmid is described in details below.

1. Delete 3 nucleotides immediately upstream of Bgl II site in pBluescript II KS-H1 vector (Brummelkamp TR et al., *Science*, 296(5567):550-553(**2002**))

PCR amplify the fragment between EcoR I-Bgl II(H1 promoter) of the original vector, with the following primers:

5' primer: GGAATTCGAACGCTGACGTCATCAACCCG

3' primer: GAAGATCTGTCTCATACAGAACTTATAAGATTCCC

(matation one: three (3) nucleotides just upstream of Bgl II site was deleted in order for transcription to start from proper position after the insertion of the AAAAA sequence according to described beneath)

Clone the PCR product in between EcoR I-Bgl II, into the original pBluescript II KS-H1 (Brummelkamp TR et al. cited above) vector, verify the plasmid DNA by sequencing:

The modified sequence:

001 TCCAGGNANC GCGGGCCCAG TGTCACTAGG CGGGAACACC CAGCGCGCGT
051 GCGCCCTGGC AGGAAGATGG CTGTGAGGGA CAGGGGAGTG GCGCCCTGCA
101 ATATTTGCAT GTCGCTATGT GTTCTGGGAA ATCACCATAA ACGTGAAATG
151 TCTTTGGATT TGGGAATCTT ATAAGTTCTG TATGAGACAG ATCTTCAATA
201 TTGGCCATTA GCCATATTAT TCATTGGTTA TATAGCATAA ATCAATATTG
251 GCTATTGGCC ATTGCATACG TTGTATCTAT ATCATAATAT GTACATTTAT
301 ATTGGCTCAT GTCCAATATG ACCGCCATGT TGGCATTGAT TATTGACTAG
351 TTATTAATAG TAATCAATTA CGGGGTCATT AGTTCATAGC CCATTATGGG
401 AGTTCCGCGT TACATAACTT ACGGTAAATG GCCCGCCTGG CTGACCGCCC

451 AACGACCCCC GCCCATTGAC GTCAATAATG ACGTATGTTC CCATAGTAAC

2. Construction of the vector with mutated dual-H1 promoters (here below referred to as pDH, stands for plasmid with Dual H1 promoters)

PCR amplify the fragment between EcoR I-Bgl II of the above modified vector, with the following primers:

5' primer: ACGCGTCGACGAATTCGAACGCTGACGTCATCAACCCG

3' primer: CCCAAGCTTGTCTCATACAGAACTTATAAGATTCCC

Clone the above PCR product in between Sal I-Hind III, in a reversed orientation, into the above modified vector, verify the plasmid DNA by Bgl II+Sal I digestion, the correct clone should have a fragment of ~1000bp. Results showed that all the ten clones checked were correct ones (Please note: pDH actually contains two truncated H1 promoter, this is due to the need to subsequent cloning process. The missing part of the promoter will be made up during the subsequent cloning process.)

3. Put the Renilla luciferase target sequence into pDH to form pDHRL: A sequence corresponding to nt 82 ~100 of Renilla luciferase mRNA was used as the test DNA. siRNA targeting this site of the Renilla luciferase was known to be active (Brummelkamp TR et al. cited above). Two oligo DNA were synthesized and annealed to each other to make the double-stranded DNA:

5' GGGGAAGATCTAAAAAAATAAATGAATCAAGAACATTTTTAAGCTTGGGG

5' CCCCAAGCTTAAAAATGTTCTTGATTCATTTATTTTTTTAGATCTTCCCC

The above double stranded DNA was cleaved with Bgl II-Hind III and cloned in between Bgl II-Hind III sites in pDH. Verification of the correct insertion of the DNA fragment into the plasmid DNA was done by cleavage by Bgl II+Sal I digestion, where the correct clone should give rise to a ~250bp fragment. All three clones tested showed to have the correct insert (Figure 2)

Efficient inhibition of luciferase expression by pDHRL. Take the above three clones: clone 1, clone 2 and clone 3, and transfect plasmid into HEK293 cells on 24-well plate, at 1.2ug, 0.6ug respectively, together with plasmid of Renilla luciferase and firefly luciferase encoding plasmids. 48 hours later, measure the Renilla and Firefly Luciferase activity. (Figure 2C). The results suggested that with the mutated promoters the plasmid can induce very efficient inhibition of the expression of target

gene Renilla luciferase, which indicated efficient production of siRNA from the mutated H 1 promoters in the dual promoter/dual terminator plasmid constructed in the current invention. Specifically the result suggested that with the mutated H 1 promoter, the RNA transcription driven by RNA polymerase III can be properly initiated and terminated, to result in the efficient production of duplex RNA of proper length that can induce significant RNA interference and inhibition of gene expression.

Cloning the randomized DNA into pDH to form a library that encodes all permutations of the siRNA

The construction of randomized DNA library that encodes all permutations of siRNA is done in a similar way as the construction of the anti-luciferase siRNA encoding plasmid in pDHRL, with the only difference that the second strand of the tester sequence was generated enzymatically to preserve the randomized nature of the sequence.

Three oligonucleotides were synthesized with 19, 20 and 21 nt of randomized region embedded within the two known sequences.

19-mer randomized region

GGGGAAGATCTAAAAA NNNNNNNNNNNNNNNNNNN TTTTTAAGCTTGGGG

20-mer randomized region

GGGGAAGATCTAAAAA NNNNNNNNNNNNNNNNNNNN TTTTTAAGCTTGGGG

21-mer randomized region

GGGGAAGATCTAAAAA NNNNNNNNNNNNNNNNNNNNN TTTTTAAGCTTGGGG

The oligonucleotides were allowed to anneal to a primer CCCCAAGCTTAAAAA and filled in with Klenow fragment in the presence of 1 mM concentration of dNTP in proper buffer (all chemicals other than DNA oligonucleotides were purchased from New England Biolabs Inc. unless otherwise specified). The duplex oligos were cleaved with Bgl II-Hind III and cloned in the Bgl II-Hind III sites of the pDH to form pDH-libraryA.

The quality of the pDH-libraryA was assessed by first clone length analysis of 41 clones, where single clone, a 10-clone pool and a 30-clone pool was used to prepared plasmid DNA and cleavage with restriction enzyme. The results suggested that all clones have the insert of the same length (figure 1B). The ten clones were

individually prepared and sequenced. All sequenced clones contain the expected 19 base pair insert as expected. Their sequences showed expected randomness as well (see below).

AAAGGGTTTACGTGGTTGG
AATCGTCTTATTTGCATGC
AATTGACATGTGAGCTTGG
AGTAGCTTGTTGAGGTTGG
CAGCATCACTGTATGTGTC
CTATCTTCGTGGAGGTTGG
CTATGAAGGTGGTGATGCG
CTTAATTGGTGGTTGTAGG
TGGCTGTATGTGAGTGGCT
TTAATCTCTGGTGTCCTAA
TTGTAGGGACTTGGATGAT

One alternative to plasmid vectors for epitopic expression of foreign gene is various types of viral vectors. Since all cloning strategies for constructing viral vectors are common knowledge, and anybody with reasonable knowledge of the art can produce viral constructs that can carry out similar expression functions as the plasmids, the disclosure of making DNA libraries as above will also enable the production of DNA libraries as such in viral vectors.

Construction of DNA libraries containing a pair of randomized regions with inverted sequences

Although the vectors with two promoters and two terminators as represented by pDHRL and *pDH-libraryA* are the preferred modes of the current invention, other methods of forming DNA libraries that encode all permutations of siRNA become obvious once the concept of DNA library encoding all permutations of siRNA is disclosed here. One such method is to form a plasmid library that encodes all permutations of the hairpin form of the siRNA. As an example, such a library can be formed according to the following procedure.

1. Library oligonucleotide was synthesized to contain a fully randomized region of 19 nt randomized sequence nested between two predetermined sequences (P1 and P2) with 5' end phosphorylated. A hairpin forming oligonucleotide was synthesized to contain 5' phosphorylation and a 3' protruding stretch

with complementary sequence of the P1 region. The Library oligonucleotide and the hairpin DNA were annealed and ligated with T4 DNA ligase and then filled in with Klenow fragment (Figure 3)

2. The extension mixture, after purification, was cleaved with BamH I and ligated into an double stranded adopter that has cohesive ends on one end and a 3' protruding stretch as a site for further priming (P3) After ligation the DNA are size-selected so that only full length fragments that contain library oligonucleotides and the hairpin oligonucleotide, as well as the adopter linker are collected.
3. Purified full length fragments are allowed to anneal to the primer 3 (which is complementary to the P3 priming site), and a strand-displacing DNA polymerase Phage29 DNA polymerase is used to drive the synthesis of a DNA fragment ALPHA. Each DNA fragment ALPHA contains: a fully double stranded adaptor linker on each end of its sequence, two identical copies of a randomized sequence arranged in reverse orientation, and the two copies are linked by the linearized sequence of the hairpin linker in its double stranded form.
4. DNA fragment ALPHA can then be cleaved in proper sites in its adoptor linker region and then ligated to a plasmid for further manipulation (Plasmid alpha).
5. Plasmid alpha is first cleaved with Sam I and Bpm I and the filled in with Klenow fragment and ligated. The resulting plasmid is propagated in E coli and then the insert is cleaved with Bcg I to remove extra sequences between the two randomized region and leave a 9-nt stretch (TTCAAGAGA) to form the loop in the future siRNA hairpin (figure 13).
6. Afterwards, the insert can be all cleaved from the plasmid with Hind III and Bgl II and inserted into the pBluescript-H1 vector to form a library. This library encodes all siRNA permutations in a hairpin form. In this case, the plasmid only need to have one promoter and one terminator for the formation of hairpin RNA within the cells.

Slight modification of the above cloning protocol as illustrated in Figure 4 can result in DNA libraries that have two wild type H1 promoter and two transcription terminators, wherein each member of the library encode the two separate strands of a double stranded RNA. This involves the insertion of a second promoter and TTTTT terminator between the two inverted randomized region of the DNA library as

illustrated in Figure 4. With the detailed disclosure described above and in figure 1-3, this alternative is obvious to a person skilled in the field.

It has to be stressed that due the enzymatic handling of the library, all siRNA that contain the restriction enzyme sites are lost. This will result in about 0.025 % siRNA loss each restriction enzyme used. So in this sense the preferred mode of the current invention, based on two promoters and two terminators, will suffer less siRNA loss and be a more complete library, than the library generated according to the above hiarpin library protocol due to the number of enzymes used in the individual protocols. Since the library contains about 2.75×10^{11} permuations in theory, the loss of siRNA species caused by the use of restriction enzymes will only have neglectable effect on the quality of library and for the screening of active siRNA against any specific gene. In the text of this invention, the referral to "all permutations of siRNA" should be understood as having this effect considered and included. Further elemination of this effect will be done by eliminating the use of restriction enzymes in the construction of the libraries.

Another note is that the sequences and restriction enzymes are only one set of examples that can be used to carry out the construction of the plasmid. The person skilled in the art can easily choose different restriction enzymes and corresponding sequences of the oligonucleotides to carry out the construction in similar manner in plasmids and viral vectors, according to the principle disclosed as above.

Generation of DNA libraries that encode cell-specific, tissue-specific or species specific double stranded RNA

With the disclosure of the random DNA libraries encoding all permutations of double stranded RNA of a given length, the method of establishing DNA libraries that encode cell-specific, tissue-specific or species specific double stranded RNA should be considered to be obvious to a person skilled in the field. One example of constructing such DNA libraries is presented below.

An oligonucleotide with 19 nt of randomized region is allowed to hybridize to mRNA purified from a specific cell type. The mRNA can be immobilized onto a streptavidin coated solid support (plastic beads for example) via biotin added to the end of the mRNA with Poly (A) polymerase. Immobilization of mRNA can be done in other ways too. After hybridization, all unbound DNA oligonucleotides are washed away and

the bound DNA sub-random oligonucleotides are collected and cloned into the vector in a protocol identical to protocols described for fully randomized DNA oligonucleotides. The libraries resulted from this process will be highly enriched for molecules that encode double stranded RNA with sequence identical to the mRNA sources.

It should be noted that although all cloning procedures herein are described in the context of a single plasmid vector, the principle should be applicable to all types of plasmids, and the cassette containing the mutated promoters, terminators and the coding region of the DNA libraries can be transferred between those different types of plasmids.

It should be further noted that although all cloning procedures are described in the context of a single type of promoter, H1 promoter, the principle should be applicable to all types of RNA polymerase III type of promoters.

One alternative to plasmid vector for epitopic expression of foreign gene is various types of viral vectors. Since all cloning strategies for constructing viral vectors are common knowledge, and anybody with reasonable knowledge of the art can produce viral constructs that can carry out similar expression functions as the plasmids, the disclosure of making DNA libraries as above should also enable the production of DNA libraries as such in viral vectors.

SUMMARY

The current invention involves DNA libraries that can generate double stranded RNA of 10 ~ 30 base pair in length, with at least one strand of the double stranded RNA having single stranded overhangs, and further involves methods to produce such DNA libraries. It is acknowledged that most frequently used double stranded RNA is siRNA of 19-21 base pair in length, normally with TT or UU overhangs on at least one of the strands. So the advantage of the current invention is discussed in comparison to siRNA generated by other methods.

In practice, only one in three to five or so short double stranded RNA that fulfill the basic structural requirement (19-21-base pair double stranded region, 3' single stranded overhangs (normally TT, or UU, but not limited to such overhangs). For knocking down the 30,000 human genes using siRNA, about 90,000 -150,000 siRNA then will have to be synthesized, at the cost of 18-30 million US dollars.

Similar amount of cost has to be allocated to any additional organism for which the full spectrum of siRNA will be generated for all genes

The current invention can generate a siRNA library encoded in plasmids that contains in theory all the permutations (4^{19}=2.75 x 10^{11}) of siRNA (19 base pair duplexes plus overhangs) (the size of libraries for double stranded RNA of other length can be easily calculated in similar way), that can be used in any organisms for which the a proper promoter(s) can be found). The cost of generating this library is just a minimal fraction of the cost of synthesizing all siRNA chemically. In other words, this is a library with the complexity of 2.75 x 10^{11} that contains reagents that can silence any gene in a mammalian and non-mammalian system. This is a very powerful toolbox for high throughput genome wide functional genomics and drug target screening, as well as nucleic acid drug development.

The complexity of this library can be further reduced dramatically by introducing a one-step oligoselection on the Library oligonucleotides. Such an approach will lead to the creation of gene-, cell/tissue-, or organism-specific siRNA encoding library that has much lower complexity (10^2-10^8), without sacrificing the usefulness of the library. Such a low complexity library can be partially or completely sequenced using different sequencing methods and enable the creation of plasmid collections that contains known siRNA encoders for each gene in an organism such as human, mouse or rat.

The description of the above is most based on plasmid system but the same library and collection can be easily established in viral vector using the same principle.

A few key classes of application of the invention is listed here as examples

1) A full collection of siRNA encoding plasmids can be selected for any given gene from this library through standard screening (which could be automated).
2) A full collection of siRNA encoding plasmids can be selected for any given cell type, tissue and organism can be established according to the invention.
3) Such collections of siRNA encoding plasmids can then be easily evaluated for their individual capacity to knockdown gene expression.
4) Most powerfully, such DNA libraries can be used for phenotype- based screening of target genes without prior knowledge of the target sequence or

the siRNA sequences , thus the artisan can avoid the biased pre-selection of target genes. This will become one most significant way of functional annotation and drug target screening.

CLAIMS

1. A DNA library for the production of a library of double stranded RNA molecules of a predefined length in the range of 10-30 base pairs in living cells, wherein the sequence(s) of the DNA region (or regions) encoding the double stranded part of double stranded RNA molecule(s) is randomized in all nucleotide positions, and wherein both strands of said double stranded RNA molecule is produced from a single member of the DNA library.

2. A DNA library for the production of a library of double stranded RNA molecules of a predefined length in the range of 19-30 base pairs in living cells, wherein the sequence(s) of the DNA region (or regions) encoding the double stranded part of double stranded RNA molecule(s) is randomized in at least 19 nucleotide positions, and wherein both strands of said double stranded RNA molecule is produced from a single member of the DNA library.

3. A DNA library for the production of a library of double stranded RNA molecules of a predefined length in the range of 15-30 base pairs in living cells, wherein the sequence(s) of the DNA region (or regions) encoding the double stranded part of double stranded RNA molecule(s) is randomized in at least 15 nucleotide positions, and wherein both strands of said double stranded RNA molecule is produced from a single member of the DNA library.

4. A DNA library for the production of a library of double stranded RNA molecules of a predefined length in the range of 10-30 base pairs in living cells, wherein the sequence(s) of the DNA region (or regions) encoding the double stranded part of double stranded RNA molecule(s) is randomized in at least 4, 7 or 10 nucleotide positions, and wherein both strands of said double stranded RNA molecule is produced from a single member of the DNA library.

5. A DNA library for the production of a library of double stranded RNA molecules of a predefined length in the range of 10-30 base pairs in living cells, wherein the sequence(s) of the DNA region (or regions) encoding the double stranded part of double stranded RNA molecule(s) is randomized in 4 to all nucleotide positions, and wherein both strands of said double stranded RNA molecule is produced from a single member of the DNA library.

6. A DNA library of any of the claims 1 to 5, wherein said double stranded RNA molecules also contain single stranded region(s) at one end or both ends of the molecules.

7. The DNA library of any of the claims 1 to 6, wherein each member of the DNA library contains one promoter for transcription of the double stranded RNA molecules and one terminator for transcription of the double stranded RNA molecules, and wherein the double stranded RNA is formed as a hairpin type double stranded molecule

8. The DNA library any of the claims 1 to 6, wherein each member of the DNA library contains at least two promoters for transcription of the components of the double stranded RNA molecules and two terminators for transcription of the components of the double stranded RNA molecules, and wherein the double stranded RNA is formed by two separate RNA molecules that are complementary to each other in the double stranded region.

9. The DNA library of claims 1 to 8, wherein the DNA library is constructed within a plasmid vector.

10. The DNA library of claims 1 to 8, wherein the DNA library is constructed within a viral vector.

11. A DNA library of claims 1 to 10 wherein the randomness of the library was modified by selection of the random DNA oligonucleotides, before cloning the said random DNA oligonucleotides into the vectors, through hybridization to a total RNA preparation or total mRNA preparation from a source, whereby only the oligonucleotides hybridized to the source RNA (or mRNA) are subsequently cloned into the vector, and wherein the source can be a cell, a cell line, a tissue, or a organism.

12. A kit containing the DNA library of any of the claims 1 to 11.

13. A method of constructing a DNA library of any of the claims 1 to 6, and 8 to 11 wherein a pair of mutated H1 promoters are placed in opposite directions to drive the RNA expression from the DNA fragment inserted between the two promoters,

wherein the said mutated H1 promoter differs from the wild type H1 promoter in at least the sequence of the 5-nucleotide region immediately ahead of the transcription starting site. The said 5-nucelotide region of the mutated H1 promoter is AAAAA.

14. An RNA library obtained from the DNA library of any of the claims 1-12, wherein the length of double stranded RNA produced is in the range of 10 to 30 nucleotides.

15. A method of using the DNA libraries of any of the claims 1 to 12, wherein the library is transiently or permanently introduced into cells as a mixture.

16. A method of using the DNA library of claims 1 to 12 to screen for double stranded RNA with biological functions.

17. A method of using the DNA library of claims 1 to 12 to screen for novel genes.

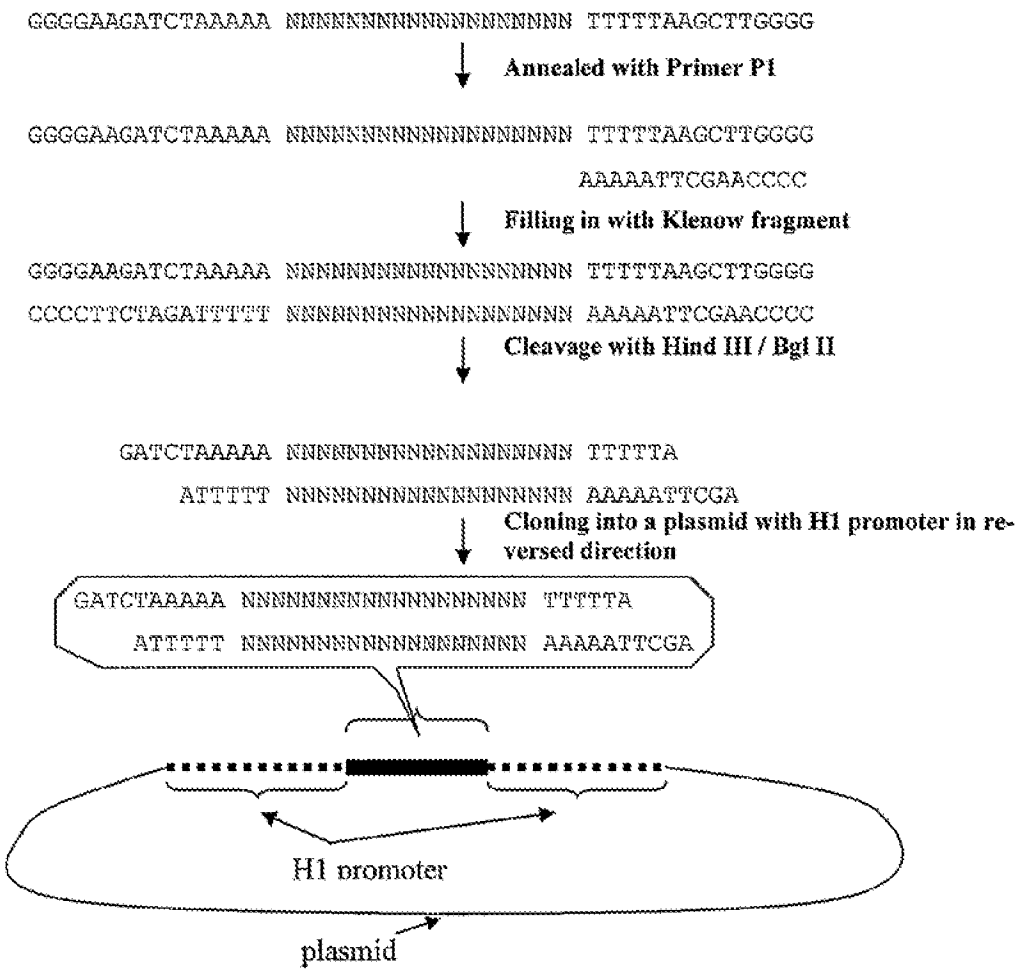
18. A novel gene obtained by the methods of any of the claims 15 to 17.

19. A novel function of a gene obtained by methods of any of the claims 15 to 17.

20.A pharmaceutical composition obtainable by the methods of any of the claims 15 to 17.

Figure 1.

A.


GGGGAAGATCTAAAAA NNNNNNNNNNNNNNNNNNN TTTTTAAGCTTGGGG
Annealed with Primer P1
GGGGAAGATCTAAAAA NNNNNNNNNNNNNNNNNNN TTTTTAAGCTTGGGG
AAAAATTCGAACCCC
Filling in with Klenow fragment
GGGGAAGATCTAAAAA NNNNNNNNNNNNNNNNNNN TTTTTAAGCTTGGGG
CCCCTTCTAGATTTTT NNNNNNNNNNNNNNNNNNN AAAAATTCGAACCCC
Cleavage with Hind III / Bgl II
GATCTAAAAA NNNNNNNNNNNNNNNNNNN TTTTTA
ATTTTT NNNNNNNNNNNNNNNNNNN AAAAATTCGA
Cloning into a plasmid with H1 promoter in reversed direction
GATCTAAAAA NNNNNNNNNNNNNNNNNNN TTTTTA
ATTTTT NNNNNNNNNNNNNNNNNNN AAAAATTCGA
H1 promoter
plasmid

B.

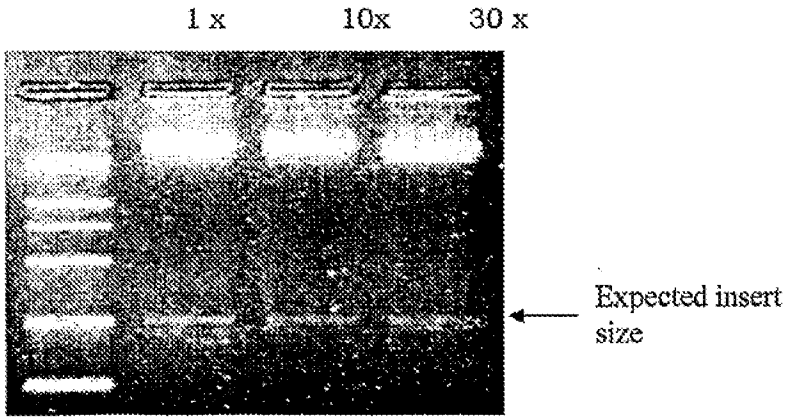
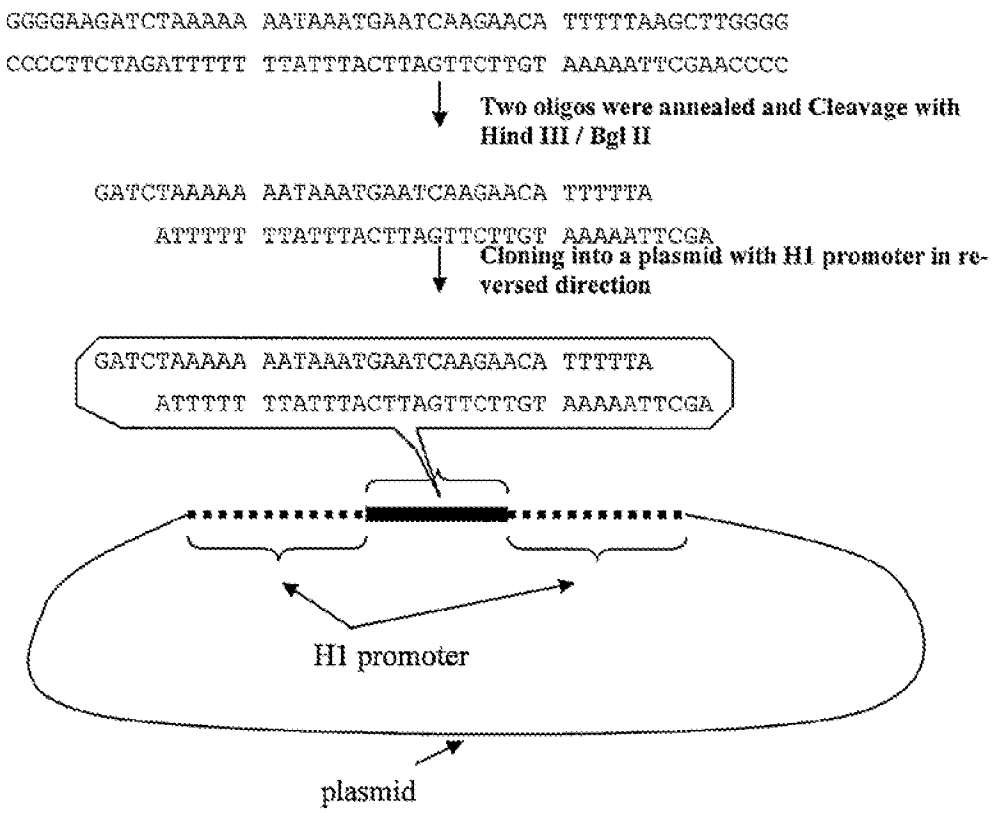

1 x
10x
30 x
Expected insert size

Figure 2.

A.


GGGGAAGATCTAAAAA AATAAATGAATCAAGAACA TTTTTAAGCTTGGGG
CCCCTTCTAGATTTTT TTATTTACTTAGTTCTTGT AAAAATTCGAACCCC
Two oligos were annealed and Cleavage with Hind III / Bgl II
GATCTAAAAA AATAAATGAATCAAGAACA TTTTTA
ATTTTT TTATTTACTTAGTTCTTGT AAAAATTCGA
Cloning into a plasmid with H1 promoter in reversed direction
GATCTAAAAA AATAAATGAATCAAGAACA TTTTTA
ATTTTT TTATTTACTTAGTTCTTGT AAAAATTCGA
H1 promoter
plasmid

B.

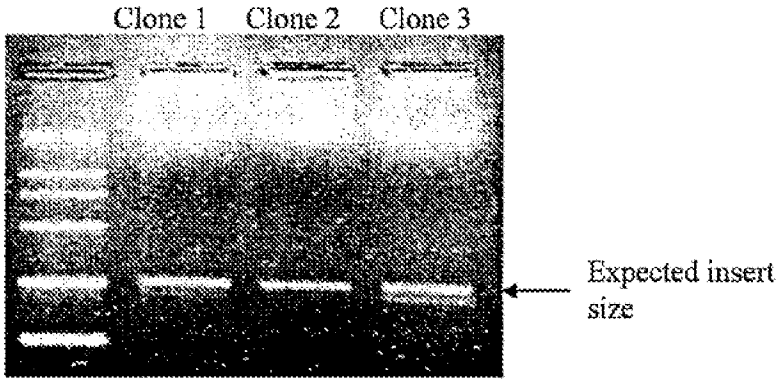

Clone 1
Clone 2
Clone 3
Expected insert size

Figure 2 (contd)

C.

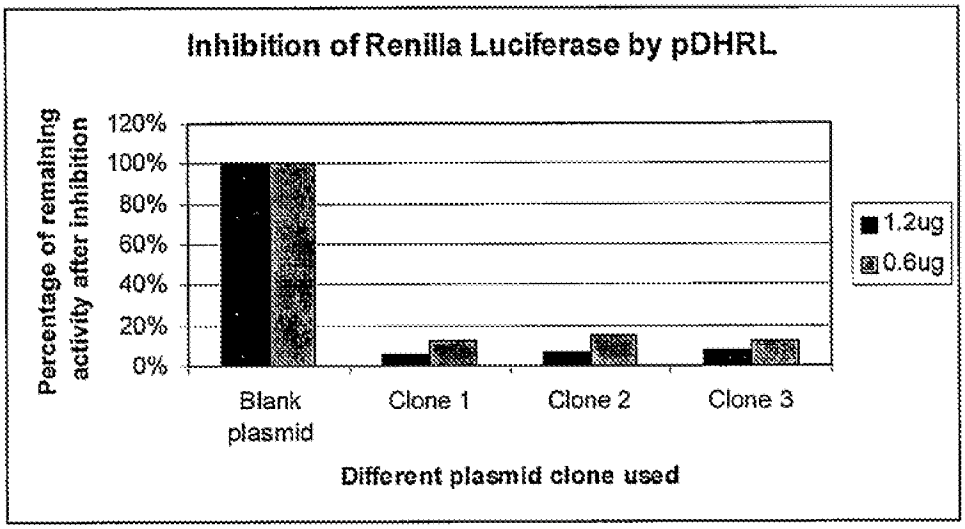
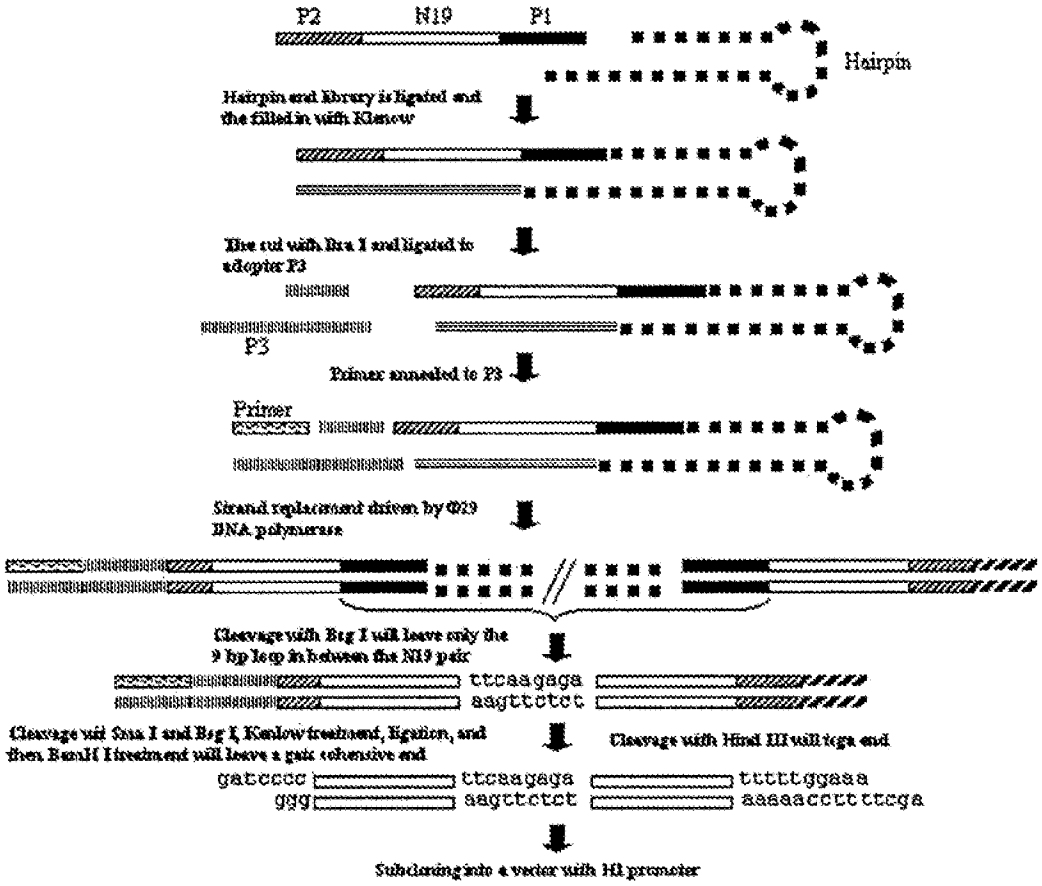


Figure 3.

A.



B.

P1: TTC AAG AGA

P2. ACA AAG CTT TTC CAA AAA

N19: NNN NNN NNN NNN NNN NNN N

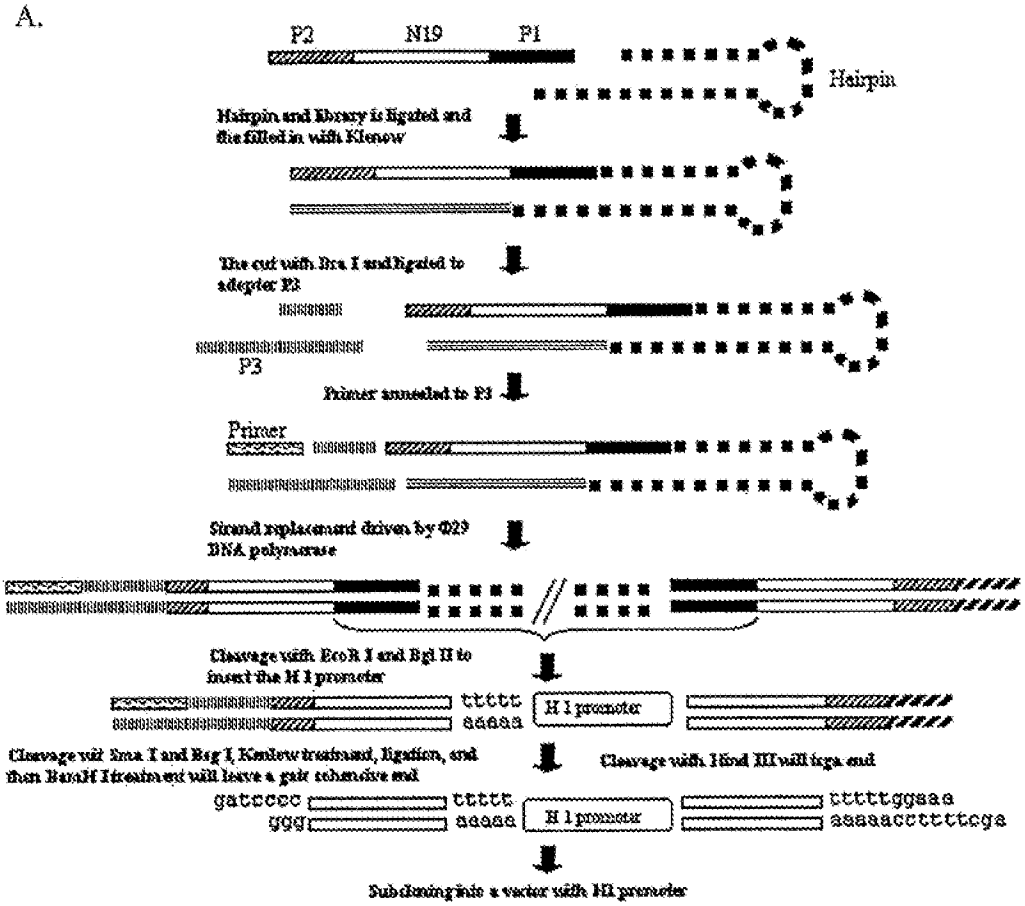
Hairpin: CAC ACG TGT CTT CGA ACA CAA TGC TAA TCT CTT GAA

P3: AGC TTA CTG CAC CC GGG GAT CCT GTT

Primer: AAC TGG ATC CCC GGG GTG CAG

Figure 4.

A.



B.

P1: TTT TTG GAT CC

P2. ACA AAG CTT TTC CAA AAA

N19: NNN NNN NNN NNN NNN NNN N

Hairpin: GGG AGA TCT TCG CTT CAA CGA AGA TCT CCC GGA TCC AAA AA

P3: AGC TTA CTG CAC CC GGG GAT CCT GTT

Primer: AAC TGG ATC CCC GGG GTG CAG

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 03/01077

A. CLASSIFICATION OF SUBJECT MATTER

IPC7: C12N 15/10, C12N 15/09, C12N 15/11, C12N 15/63
According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

IPC7: C12N, C12Q

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

SE,DK,FI,NO classes as above

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

EPO-INTERNAL, WPI DATA, PAJ, BIOSIS, MEDLINE

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Nature Biotechnology, Volume 20, 2002, Makoto Miyagishi et al, "U6 promoter-driven siRNAs with four uridine 3' overhangs efficiently suppress targeted gene expression in mammmalian cells", pages 497-500, abstract, figure 1, page 499, left column, paragraph 2	1-17
	--	
X	WO 0222634 A1 (SANGAMO BIOSCIENCES, INC.), 21 March 2002 (21.03.02), abstract, page 8, lines 13-15, page 30, line 3 - page 34, line 20, page 35, claim 1, page 38, claims 16-17	1-6,9-10,12, 14-15
	--	
X	WO 0175178 A2 (ENANTA PHARMACEUTICALS, INC.), 11 October 2001 (11.10.01), page 8, line 7 - line 25; page 12, line 8 - line 10	1-5,9-10,12, 15
	--	

[X] Further documents are listed in the continuation of Box C. [X] See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be of particular relevance
"E" earlier application or patent but published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance: the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance: the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
30 Sept 2003	07 -10- 2003
Name and mailing address of the ISA/ Swedish Patent Office Box 5055, S-102 42 STOCKHOLM Facsimile No. +46 8 666 02 86	Authorized officer SARA NILSSON/BS Telephone No. +46 8 782 25 00

Form PCT/ISA/210 (second sheet) (July 1998)

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 03/01077

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	US 5824469 A (HORWITZ ET AL), 20 October 1998 (20.10.98), abstract, column 7, line 57-67, column 29, claim 2	1-5,9-10,12, 15
A	National Library of Medicine, (NLM), file Medline, Medline accession no. 12903211, Kawasaki Hiroaki et al, "A functional gene discovery in cell differentiation by hybrid ribozyme and siRNA libraries" & Nucleic acids research. Supplement (2001), England 2002, No. 2, pages 275-276, last 7 lines	1-17
A	WO 0005415 A1 (IMMUSOL INCORPORATED), 3 February 2000 (03.02.00), page 18, line 20 - line 27; page 19, line 21 - line 32; page 24, line 29 - page 25, line 3, figure 1	1-17
A	WO 9961631 A1 (NOVARTIS AG), 2 December 1999 (02.12.99), page 13, paragraph 2	1-17
A	US 6153380 A (NOLAN ET AL), 28 November 2000 (28.11.00), column 3, line 58 - line 60; column 13, line 16 - line 48; column 101, claim 1	1-17
A	WO 0188121 A1 (DEVGEN NV), 22 November 2001 (22.11.01), see especially abstract, page 3, line 26 - page 4, line 4	1-17
A	PNAS, Volume 99, No. 8, 2002, Guangchao Sui et al, "A DNA vector-based RNAi technology to suppress gene expression in mammalian cells", pages 5515-5520, abstract, page 5515, left column, paragraph 2 - right column, paragraph 1, page 5517, figure 1A	7

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 03/01077

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT

Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
A	Molecular and Biochemical Parasitology, Volume 111, 2000, Douglas J. et al, "Double-stranded RNA interference in Trypanosoma brucei using head-to-head promoters", pages 67-76, abstract, page 68, right column, lines 7-13, pages 74, right column, paragraph 3 --	1-17
A	WO 0001846 A2 (DEVGEN N.V.), 13 January 2000 (13.01.00), abstract --	1-17
P,X	EP 1229134 A2 (NUCLEONICS, INC.), 7 August 2002 (07.08.02), column 1, paragraph [0004], column 3, line 45, column 16, paragraph [0057], column 17, lines 19-20, column 24, line 53 - column 25, line 2 --	1-17

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE03/01077

Box I Observations where certain claims were found unsearchable (Continuation of item 1 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:
because they relate to subject matter not required to be searched by this Authority, namely:

2. ☒ Claims Nos.: 18-20
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:

see next sheet

3. ☐ Claims Nos.:
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

Box II Observations where unity of invention is lacking (Continuation of item 2 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:

4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐ The additional search fees were accompanied by the applicant's protest.

☐ No protest accompanied the payment of additional search fees.

Present claims 18-20 relate to a gene, a novel function of a gene or a pharmaceutical composition defined by being obtained by the method according to claims 15-17. The claims cover all genes, functions and compositions having this characteristic, whereas the application provides support within the meaning of Article 6 PCT and disclosure within the meaning of Article 5 PCT for no such genes, functions or compositions. Additionally, previously known genes, functions or compositions may be included in the scope of the present claims. In the present case, the claims so lack support, and the application so lacks disclosure, that a meaningful search over the whole of the claimed scope is impossible. Independent of the above reasoning, the claims also lack clarity (Article 6 PCT). An attempt is made to define the genes, functions and compositions by reference to a result to be achieved. Again, this lack of clarity in the present case is such as to render a meaningful search over the whole of the claimed scope impossible. Therefore, no search has been performed for claims 18-20.

The applicant's attention is drawn to the fact that claims, or parts of claims, relating to inventions in respect of which no international search report has been established will not be the subject of an international preliminary examination (Rule 66.1(e) PCT). This is the case irrespective of whether or not the claims are amended following receipt of the search report or during any Chapter II procedure.

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 03/01077

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0222634 A1	21/03/02	AU 8608301 A	26/03/02
		GB 0022330 D	00/00/00
WO 0175178 A2	11/10/01	AU 8928401 A	15/10/01
		US 2003108532 A	12/06/03
US 5824469 A	20/10/98	NONE	
WO 0005415 A1	03/02/00	AU 750598 B	25/07/02
		AU 5117399 A	14/02/00
		CA 2335390 A	03/02/00
		EP 1097244 A	09/05/01
		JP 2002528049 T	03/09/02
		US 2003096399 A	22/05/03
WO 9961631 A1	02/12/99	AU 751409 B	15/08/02
		AU 4368399 A	13/12/99
		BR 9910729 A	13/02/01
		CA 2328058 A	02/12/99
		CN 1307641 T	08/08/01
		EP 1080208 A	07/03/01
		HU 0102103 A	28/09/01
		IL 139374 D	00/00/00
		JP 2002516112 T	04/06/02
		PL 344312 A	22/10/01
		TR 200003481 T	00/00/00
		ZA 200006594 A	14/11/01
US 6153380 A	28/11/00	US 2002127564 A	12/09/02
		US 2002146710 A	10/10/02
		US 2003104384 A	05/06/03
		AU 725716 B,C	19/10/00
		AU 1707497 A	20/08/97
		AU 1707897 A	20/08/97
		CA 2244222 A	31/07/97
		DE 69722176 D	00/00/00
		EP 0877752 A,B	18/11/98
		EP 1295952 A	26/03/03
		JP 2000504220 T	11/04/00
		JP 2001509661 T	24/07/01
		US 6365344 B	02/04/02
		US 6455247 B	24/09/02
		US 2003044767 A	06/03/03
		WO 9727212 A	31/07/97
		WO 9727213 A	31/07/97
WO 0188121 A1	22/11/01	AU 6416601 A	26/11/01
		EP 1285062 A	26/02/03
		GB 0012233 D	00/00/00
		GB 0112213 D	00/00/00
		GB 2365867 A,B	27/02/02

INTERNATIONAL SEARCH REPORT

International application No.
PCT/SE 03/01077

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0001846 A2	13/01/00	AU 4907999 A	24/01/00
		BR 9911802 A	22/01/02
		CA 2332619 A	13/01/00
		CN 1323354 T	21/11/01
		DE 1093526 T	11/10/01
		DE 29924298 U	12/09/02
		DE 29924299 U	12/09/02
		EP 1093526 A	25/04/01
		EP 1197567 A	17/04/02
		GB 0020485 D	00/00/00
		GB 0118514 D	00/00/00
		GB 0206600 D	00/00/00
		GB 2349885 A,B	15/11/00
		GB 2362885 A,B	05/12/01
		GB 2370275 A,B	26/06/02
		GB 9814536 D	00/00/00
		GB 9827152 D	00/00/00
		HU 0103571 A	28/01/02
		IL 140467 D	00/00/00
		JP 2002519072 T	02/07/02
		NO 20010019 A	05/03/01
		PL 347978 A	06/05/02
		US 2003061626 A	27/03/03
		ZA 200007653 A	19/09/02
EP 1229134 A2	07/08/02	AU 1473502 A	01/08/02
		CA 2369944 A	31/07/02
		US 2002132257 A	19/09/02